December, 2004

County of Mono, Ca

Geospatial Data Standards and Contractor Guidelines

Developed by:
Nate Greenberg
GIS Coordinator
County of Mono, Ca
PO Box 347
Mammoth Lakes, Ca 93546
(760) 924-1819
ngreenberg@mono.ca.gov

MONO COUNTY GIS

Geospatial Data Standards and Contractor Guidelines

INTRODUCTION

Background:

The goal of Mono County Geographic Services is to maintain a system which integrates a wide range of data sets from a variety of different users, departments and outside sources. Through GIS Technology, multiple datasets may be integrated, tying one piece of information to another through commonalities including geographic location and feature descriptions.

While GIS in Mono County is still in its infancy, a great deal of effort has been made to create a standardized, stable platform for current and future users. The core of this platform resides on a clear definition of system configuration and design and how applications and users interact with the integrated data.

Purpose of Document:

This document is designed to serve both as background information and direct guidelines for the development, use and integration of geospatial data as it relates to the County's Geographic Information System. This document is not intended to provide detailed steps in the creation, use or integration of new data, but to orient users with the expectations of accuracy, organization and documentation for all data related to this system.

In addition, it is expected that this document accompany all Request for Proposal (RFP) documents, Scope of Work Documents and Contractual Agreements between the County and any contractor working on projects that involve the use or creation of geospatial data. Approval of a Scope of Work or Contractual Agreement also assumes approval of the terms, conditions and standards established in this document.

Definition of Terms:

CAD: CAD stands for Computer Aided Design, which is The use of computer programs and systems to design detailed two- or three-dimensional models of physical objects, such as mechanical parts, buildings, and molecules.

Corporate Data Library: All geospatial and related data within the Town is stored on a server in a folder structure that keeps the information organized.

Digitizing: The process of converting non-digital formats of information, such as paper maps, into digital information in the computer. In this case, digitizing refers to creating geospatial data from another source of information.

GIS: Geographic Information System, the collection of hardware, software, databases and geospatial data that enable the reception, manipulation, analysis and display of information pertaining to a landscape or geographic area.

Geodatabase: The Geodatabase (short for geographic database), defines all the types of data that can be used in ArcGIS – including, features, rasters, addresses, and survey measurements — and how they are represented, accessed, stored, managed and processed.

Geospatial Data: The digital, geographic and location-specific information including related attribute records, data files and metadata that are used within a GIS.

License Agreement: A legal document detailing the terms and conditions associated with the use of the agency's geospatial data.

Licensee: Any recipient of the agency's data holdings including the public, other governmental agencies or contractors and consultants hired by the agency.

Metadata: Information and documentation that describes geospatial data including contact, accuracy, development protocols and standards, process, coordinate system, currency, and format information.

Shapefile: A shapefile stores nontopological geometry and attribute information for the spatial features in a data set. The geometry for a feature is stored as a shape comprising a set of vector coordinates.

Topology: The mathematical concept of Leonhard Euler, which describes the spatial relationships between neighboring or adjacent features such as points, lines and polygons.

Assumptions:

- All individuals utilizing the agency's geospatial data are familiar with the concepts and applications of GIS.
- Users and developers of geospatial data are comfortable with concepts of data organization, management and maintenance, including documentation.
- Developers of GIS data are well versed in data management and organization, including issues associated with accuracy, process documentation, naming conventions, and the standards addressed in this document.

Licenses and Data Sharing Agreements:

All geospatial data maintained by the County is protected under our Geospatial Data License Agreement. Use of the data is contingent upon the acceptance of this agreement and the terms provided within it. Under the agreement, no data shall be re-issued or re-distributed by any Licensee of the data. All requests for data must be met by the Licensing agency, unless a separate contractual agreement between the agency and Licensee exists.

The biggest implication of this data license process is that no user of the County's geospatial data possesses the rights to that data, and therefore is not entitled to distribute it as his or her own. If subcontractors are used during a project, approval must be obtained before the data can be passed along, and then may only be done so in conjunction with the acceptance of the Geospatial Data License Agreement by the subcontractors. For more information, please refer to the License Agreement or contact the County's GIS Coordinator.

DATA BACKGROUND

Data Organization and the Corporate Data Library:

The Corporate Data Library contains geospatial data that can be read into various GIS applications, including Desktop and Web-based. This library resides on the GIS Server (\\Mapserver1) and can be accessed by any user on the County's network. The library is a "One-Way Access" library, which means that any user can read from it, but only administrators can write to it. This is to ensure that source data sets are not tampered with or corrupted during mapping or analysis projects.

Corporate datasets are created by departmental GIS users, consultants, or are acquired from outside sources (such as other agencies). Typically these data sets describe features and conditions that benefit multiple users or are key pieces of 'source' information from which decisions can be made. When changes and updates happen to these data sets, they are well documented with a description of the action and the time at which it happened. Often, the history of pre-existing conditions is maintained by backing up previous versions of the data sets.

All data sets in the library are properly indexed and described with Metadata which can be reviewed by the user through created HTML or XML documents. Metadata is a vital piece of

information for all data sets in the library because it describes the data content, its accuracy and how current it is.

Data Format:
Mono County maintains geospatial data in both ESRI Shapefile and Geodatabase formats. All data editing and management is carried out in the Geodatabase data structure for ease in management and advanced editing and topological tools. The Geodatabase (short for geographic database), defines all the types of data that can be used in ArcGIS – including, features, rasters, addresses, and survey measurements—and how they are represented, accessed, stored, managed and processed. Once data sets are complete, they are exported to Shapefiles for users to bring into desktop GIS applications or for distribution to other users and contractors outside the County.

All users will work within the Shapefile data format. If new data sets are developed through a project, they are to be delivered back to the County in Shapefile format. In most cases, this data will be integrated into the Geodatabase for management purposes.

Data Standards:
Standardization in the creation of GIS data is incredibly important as it determines the success or failure in terms of relationships with other data sets. While there are many different standardization factors that need to be met, there are three in particular which the County of Mammoth Lakes GIS strives to maintain.

Projection / Coordinate System:
Any new data that is created for the County must be created in our coordinate system. In addition, all datasets must have an ArcView Projection File (.prj) accompanying them that references the data set to that coordinate system. The coordinate system we use is:

California State Plane, Zone III, NAD 83, US Feet

Database and Attributes:
Every feature within a data set must be populated with a set of attributes. There should be no NULL values anywhere in the table. In addition, the dataset must contain a unique identifier. This ID can be a number of different things, from an APN to a uniquely auto-assigned ID. All attribution must remain consistent in data entry. Here are just a few thoughts and examples related to consistency of attributes:

- Only one type of value should be entered into each database column, for example, address should be broken up into (1) House Number, (2) Street Name, (3) City, (4) State, and (5) Zipcode.
- People's names should be broken up by First and Last
- APNs should be entered in their full 15-digit format, without hyphens
- Description fields are fine, but should be coupled with a unique ID for each feature

Metadata:
All data sets in the County's GIS Library must contain complete Metadata. Metadata upkeep is very important, because whenever data sets leave the County and are utilized in other GIS systems, the operators of those systems need to know how to use the data properly.

The County maintains all Metadata via the Metadata Editor in ArcCatalog. If you have not used the metadata editor in ArcCatalog, refer to Appendix A. If you do not have

access to ArcGIS (but are creating new datasets), please refer to and complete the Metadata template in Appendix B for each new dataset.

Fields that must be populated in ArcCatalog Metadata are as follows:

a. Abstract: What is the data about? Write a brief 1-4 sentence statement.
b. Purpose: Beyond the Abstract, what is the purpose or need for the data set? (1-2 sentences)
c. Use Constraints: Are there restrictions or limitations associated with the data set? Can it be distributed, etc?
d. Contact: Fill out all applicable information for the individual / organization that developed the data.
e. Coordinate System: While all data should be created in California State Plane, Zone 3, NAD 83, US Feet, the Coordinate System information should nevertheless be filled in.
f. Status: Determine whether the data set is complete, in work, etc. and the note the publication date.
g. Source: Where did the data originate from? Include the entity, individual and creation date.
h. Keywords: Populate at least one or two keywords for each dataset. Please refer to the Federal Geospatial Data Committee keyword standards at www.fgdc.gov for more information.
i. Completeness Report: State whether or not the data is complete, and, if not, how and where it is incomplete
j. Accuracy Report: Explain in 1-2 sentences the accuracy of the data set, pertaining to positional accuracy (matching reference layers, etc), and any known inaccuracies in either attribute or feature data.
k. Attribute Definitions: Fill out the 'Definition' field for each of the attribute fields in your data set. Any fields with common codes, such as Zoning or Land Use should have definitions for those codes as well.

Accuracy and Consistency:

Building accurate data sets from scratch is not an easy task and requires extensive knowledge and experience with GIS. When creating new data sets, it is important to consider the two areas where inaccuracies could be introduced; (1) conversion from the source information to digital format (digitization, etc.) and (2) topological development and related errors.

When creating new data from non-digital sources, it is important to consider the different methods that can be used for bringing this data in. Tablet-based or heads-up digitizing are the two most common approaches to this, and there are advantages and disadvantages with each that must be understood prior to use. Regardless of the method you use, it is important to document the process in the Accuracy Report, as well as the level of error that you believe to exist in the data. The same documentation criteria exist for integrating existing digital data, such as non-georeferenced CAD or GIS data. Make sure that you evaluate the positional accuracy of your adjusted data.

Topological errors associated with data creation typically are related to the relationships between features within the same dataset, or supporting data sets. When creating new data that relates to other existing data, it is highly recommended that the base data be used as the starting point for this effort. For example, when creating Zoning Blocks, which are based on existing parcels, the parcel data should be used as the base. Instead of re-creating line or polygon segments that represent the boundaries of these blocks, the actual parcel edges should be used to define these boundaries. By doing so, you can be assured that the topologies of the two layers will match and that related

topological errors can be avoided. If you have questions about the creation of new data relative to existing data, please contact the GIS Coordinator.

File Naming Conventions:
How created data sets are named is also important, particularly for future recognition and use. Currently, the Mono County GIS Library is broken up into several categorical folders that contain data sets. These folders include;

- Aspatial: This folder contains databases, such as the Assessor Database that can be connected to other data sets in the Library.

- Boundaries: This folder contains data sets that represent regions such as the Mono County Boundary, gross-scale land ownership, USGS quad boundaries, etc.

- Cartography: This folder contains data that is build specifically for use in mapping and contains information such as place names and annotation.

- Environmental: This folder contains data sets that relate to environmental conditions such as Vegetation cover, Soils and the locations of Special Status Species.

- Health: This folder contains data related to the Mono County Health Department.

- Imagery: This folder contains 'basemap' type images such as Topo maps and Aerial Photographs which other data can be overlaid on.

- Parcels: This folder contains the Mono County Parcel Basemap and other related data such as locations of Fire Hydrants and Building Footprints.

- Pictures: This folder is intended to be a resource of photographs for various structures and locations throughout the county that can be related to GIS.

- Planning: This folder contains data more specifically related to the Planning Department including Special Districts, Planning Areas, Supervisorial Districts, etc.

- Transportation: This folder contains data sets such as Highways and Minor Roads, as well as CalTrans generated GIS data sets.

- Water: This folder contains data sets related to hydrology, such as Streams, Lakes and Watersheds..

All data sets should be named in a 'DOS' friendly manner, which means names CAN NOT CONTAIN spaces, special characters (such as !,@,#,$,%,^,&,*,(),:,;,",',<,>,.,,,?,/,etc.). If the data set requires two names to describe it, which is often the case, please utilize either a hyphen (-) or an underscore (_) to separate the names, for example "mono_parcels" or "assessor-database".

The only time that a convention exists for naming is related to incorporating a date into the name, such as for Parcels that were created on December 3, 2004. In this case the month, day and year of the date are lumped together – year, month, day, and the entire date is separated from the rest of the name with an underscore like: "Mono-Parcels_20041203".

DATA USE

License Agreement / Terms of Use:

Please refer to the complete License Agreement for a more complete definition of the permitted and non-permitted uses of the data.

Metadata Review:

Before using any data set that you are unfamiliar with, it is highly recommended that you review the associated Metadata for data Purpose and Limitations. It is often valuable to review the Attribute definitions as well so that you know exactly what you can and can not get out of the data.

Support:

If there are any questions pertaining to the data that cannot be answered by the Metadata, please contact the County's GIS Coordinator, Nate Greenberg, before using the dataset.

DATA DELIVERY / INTEGRATION

Organization and Cataloging:

All data delivered to the County upon completion of a project must meet the criteria described above in the Data Standards section. For your convenience, a checklist can be found in Appendix C.

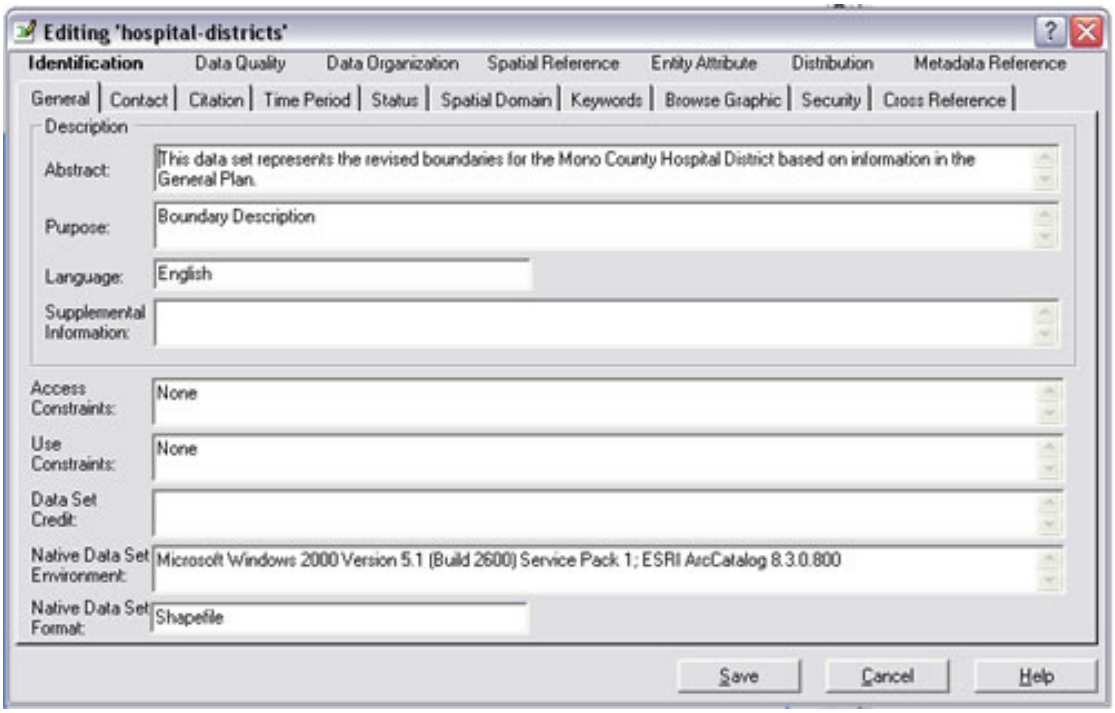
Data should be organized and delivered on CD or DVD media and broken up into thematically named folders, if multiple data sets are present. A 'Readme' file in .txt format should be written and placed in the root of the CD or DVD describing what the organization structure of the media is. If projects, digital maps and data are all contained on the media, please create a folder for each (eg. \data\, \projects\, \products\), in addition to creating subfolders under these, if needed. The goal is to allow the GIS technician to be able to open all the projects from the CD without having to relocate data. In order to accommodate this, please ensure that all map documents are stored with relative paths.

Appendix A: Creating Metadata with ArcCatalog

Below are instructions for completing these sections of the Metadata, which should be followed once your data is completed and has been saved.

1. Open ArcCatalog from ArcMap or your start menu and navigate to the directory where your data set is saved
2. Select the dataset in the left view pane 

then select the 'Metadata' folder tab.
3. The existing metadata for the dataset will appear in the right view pane – it should be blank for the most part.
4. To start editing the Metadata, click on the Write button  which will open the Metadata Editor used to populate necessary fields.



5. The information that needs to be populated is as follows:
    a. Identification > General > Abstract: What is the data about. Write a brief 1-4 sentence statement
    b. Identification > General > Purpose: Beyond the Abstract, what is the purpose or need for the data set (1-2 sentences)
    c. Identification > General > Use Constraints: Are there restrictions or limitations associated with the data set. Can it be distributed, etc?
    d. Identification > Contact: Fill out all applicable information for the individual / organization that developed the data.
    e. Identification > Status: Determine whether the data set is complete, in work, etc.
    f. Identification > Keywords: Populate at least one or two keywords for each dataset.
    g. Data Quality > General > Completeness Report: State whether or not the data is complete, and if not, how and where it is incomplete
    h. Data Quality > Attribute Accuracy > Accuracy Report: Explain in 1-2 sentences the accuracy of the data set, pertaining to positional accuracy

(matching reference layers, etc), and any know inaccuracies in either attribute or feature data.

i. Entity Attribute > Detailed Description > Attribute: Fill out the 'Definition' field for each of the attribute fields in your data set. You can use the Forward and Back arrows at the bottom of the pane to scroll between the different attributes.

6. Once you have entered values for all above mentioned areas, click the Save button. You have now created Metadata for your dataset.

Appendix B: FGDC Metadata Template

DATE:
COVERAGE NAME: <zzzzzzzzzzz>.SHP
METADATA PREPARER:
- Name:
- Company
- Address
- Phone
- Email:

ABSTRACT:
DATA DESCRIPTION: (ex. This coverage depicts…)
GEOGRAHIC EXTENT: <State, Ecoregion, Site>
KEY WORDS:
COORDINATE SYSTEM DEFINITION:
- Projection:
- Zone:
- Units:
- Datum:
- Spheroid:
- 1st Std. Parallel:
- 2nd Std. Parallel:
- Central Meridian
- Longitude of Origin:
- Latitude of Origin:
- False Easting (X shift):
- False Northing (Y shift):

SOURCE(s): (ex. USGS 1:24,000 dlg's)
SOURCE PROJECTION:
DATE UPDATED
UPDATE BY:
NOTES / COMMENTS:
ENTITY AND ATTRIBUTE INFORMATION:
- Field Name:
- Field Description:
- Value Definitions:

PROCESS STEP LINEAGE: Date and step. Can get from log file if using ArcInfo
NOTES:
DATA QUALITY ASSESSMENT:

Appendix C: Data Development and Delivery Check List

Standardization

___ All features in data set have populated attributes.

___ A unique identifier exists for every feature in the dataset.

___ All attribute values within each field are formatted identically (eg. Land Use Codes, etc).

___ Metadata is complete for the dataset based on the requirements presented in the above sections.

___ Metadata is stored as .xml via ArcCatalog or is stored as a .doc file using the template in Appendix B, and saved with the same name as the dataset.

___ The data was created using the most accurate input methods possible.

___ An accuracy assessment was completed for the dataset, documenting the creation process.

___ Topological considerations were met and no topological errors exist.

___ Naming conventions are followed (no spaces or special characters).

___ The name of the data set clearly describes what the data represents.

___ A category recommendation is made for where the data should reside in the Corporate Library.

Delivery and Integration

___ Data is organized into folders and sub-folders within the delivered media.

___ Projects, Data and Products are separated into folders and sub-folders.

___ Relative paths are used in all included project files. Projects will open directly from CD or DVD.

___ A 'Readme' file exists in the root of the delivered media describing the included folders with contact information for the delivering party.